Referring to Fig. 2 there is now shown a conceptual diagram of how the predictive model detects premium fraud. The collection of insurance policies upon which the predictive model is developed form a complex multi-dimensional "policy space" 201, which contains all of the policies that will be evaluated by the predictive model. Each policy is described by many policy variables. These policy variables generally fall into three categories of variables: over-time policy variables 203, peer group variables 205, and internal policy variables 207. It is this collection of policy variables that describes each policy in the policy space 201. In general, many of these variables may be understood as measures of the amount, distribution, or nature of the activities or characteristics of the policyholder and its claimants as indicators of premium fraud risk.

IN THE DETAILED DESCRIPTION:

Please replace the paragraph beginning at line 20 of page 42 with the following:

4. Policies with an officer who is currently or was recently an officer on a different policy and where the new policy has a lower experience modification rate than the previous policy. The logic here attempts to identify policies that may be evading high experience modification rates by closing the company and re-opening it under a new name.

5. Policies that have a class code on a claim for which no premium was reported at the time the claim was opened. The logic here is similar to the first rule, except in this case the job class code is listed on the payroll report but no payroll is reported in that class code. This may imply that the employer is misrepresenting the job classifications of their payroll in order to lower their premium.

Each rule in the rule-based analysis 620 flags any policies that violate the rule. These flags can be used to create lists of violators, which are useful complements to the scores from the predictive model 622. As noted above, in a workers' compensation implementation, policies with zero payroll are not scored by the predictive model 622, so without the rule-based analysis, suspicious policies in that group would not be evaluated. While the exclusion of such policies from the predictive model 622 is appropriate, it may still be possible to identify suspicious policies in this group, as the above rules demonstrate. Thus, the rule-based analysis 620 provides such analysis, bringing any problem policies with zero payroll to the attention of auditors. The rule-based analysis can also provide valuable additional analysis for policies that are scored by the predictive model 622. For example, a policy with a class code on a claim that is not on the policy might be scored by the predictive model 622, but if nothing else about that policy looks suspicious, it may not score high. The rule-based analysis 620 however would flag such a policy as having a clear-cut, specific problem that is independent of how suspicious the policy looks more generally.

Please replace the paragraph beginning at line 2 of page 68 with the following:

A randomly selected portion (*e.g.*, 20-30%) of the model development dataset is held out 909 from model training. This hold-out set is referred to as the "test" data 908b and is used to test the model that is trained on the remaining dataset 908a portion of the dataset 906a. Evaluation of the hold-out data ensures that the predictive model 622 does not over-fit the training data 908a. Also, the test data can be used to estimate the production performance of the model (indeed, of the entire system).

<u>IN THE ABSTRACT:</u>

Please replace the Abstract with the following paragraph:

Detection of insurance premium fraud is provided by a predictive model, which uses derived variables to assess the likelihood of fraud for each policy. The predictive